A Hybrid Multi-stage Network for Lung Segmentation, Disease Classification and Severity Localization from X-ray Images

Shubham Garg (sg8311), Somya Gupta (sg7885), Banani Ghosh (bg2502)
New York University

Introduction

Accurate and rapid diagnosis of respiratory diseases such as COVID-19 and viral pneumonia using chest X-rays (CXRs) is crucial for timely treatment and containment efforts. However, traditional diagnostic approaches often struggle with high variability in image quality and the subtlety of disease manifestations, leading to a significant rate of diagnostic errors. To address these challenges, this report presents a hybrid multi-stage network that initially segments the lung region in the CXR images, followed by classification and subsequent localization of the disease using Grad-CAM. This approach allows for focused analysis on relevant lung areas, enhancing the model's accuracy and reliability in diagnosing respiratory conditions.

Literature Review

Recent advances in deep learning have significantly impacted medical imaging, particularly in the automated analysis of chest X-ray images for disease detection and classification. Segmentation of chest X-rays using deep learning models like U-Net, SegNet, and LinkNet has demonstrated considerable success in isolating regions of interest such as lung fields, which is critical for accurate disease diagnosis [2]. However, these models often require substantial datasets for training to achieve high accuracy and generalizability.

The application of deep learning extends to the classification of lung diseases using various network architectures. Notably, the integration of models like Vision Transformer and ResNet50 [8] into the disease classification process has shown promising results, leveraging their capabilities to understand complex patterns within medical images [1]. Techniques like GradCAM further enhance model interpretability by providing visual explanations for the decisions made by the network, crucial for validating AI assessments in clinical practice [2].

Despite these technological advances, challenges remain, such as data scarcity and the high variability in medical images due to different acquisition protocols. Furthermore, issues related to the training volume, as highlighted in recent studies, show that the performance of deep learning models like the Vision Transformer depends heavily on large-scale pre-training datasets which are often not available in medical imaging [1]. Additionally, models trained on general datasets might not transfer well to medical applications due to distinct image characteristics, underscoring the need for domain-specific tuning [6].

The variability of imaging conditions and the domain-specific nature of medical datasets necessitate tailored approaches, as generalization across different institutions presents a significant hurdle [4]. Furthermore, the sensitivity and specificity of these models can significantly fluctuate based on the dataset and the particular characteristics of the diseases being analyzed, with recent works proposing hybrid models combining CNNs and transformers to improve robustness [5].

In conclusion, while deep learning has significantly advanced the field of lung segmentation and disease classification in X-ray images, the robustness and applicability of these models in clinical settings continue to face challenges. Addressing these will require ongoing research aimed at enhancing data handling, model interpretability, and generalization across diverse clinical environments.

Methodology

Overview

Our approach involves a systematic progression through three stages: segmentation, classification, and localization. This structured workflow allows us to precisely isolate and analyze lung regions, identify pathological conditions, and visually highlight critical areas influencing diagnostic outcomes, thereby facilitating a comprehensive examination of CXR images.

Segmentation

- **Problem being addressed:** Much of the prior research has focused on extracting features from entire CXR images. However, irrelevant features for diagnosis, such as dark patches, parts of the patient's body outside the lungs, and lung margins, can negatively affect the decision-making process and increase computational costs. It is crucial to remove these non-essential features to allow classifiers to focus exclusively on the specific pathological areas within the CXR images.
- **Models Tested:** We explored several advanced segmentation models to accurately delineate lung regions in CXRs.

These included U-Net, FPN, and LinkNet, each tested with multiple backbones such as VGG16, SEResNeXt50, InceptionResNetV2, MobileNetV2, and EfficientNetB2.

- **Selection:** The ensemble model, combining Inception-UNet, SEResNeXt50-LinkNet, and EfficientB2-FPN, was selected for its superior segmentation capabilities, particularly for lung areas. This approach leverages the strengths of each architecture: Inception-UNet for capturing intricate spatial hierarchies, SEResNeXt50-LinkNet for its ability to focus on crucial features through residual connections and attention mechanisms, and EfficientB2-FPN for its efficient scaling and feature pyramid networking.

  The ensemble strategy is particularly effective at maximizing the segmented image area, significantly reducing information loss during the segmentation process. By pooling the unique strengths of each model, the ensemble ensures that crucial features are retained and enhanced in the final segmented output. This method not only improves the precision of segmentation but also ensures that the subsequent analysis is based on the most comprehensive data available from the imaging. The use of multiple models in an ensemble also compensates for any individual model's weaknesses, providing a more reliable and detailed capture of lung regions, crucial for accurate biomedical segmentation.

- **Hyperparameters:** The model was set up with an input shape of 256×256 pixels for grayscale images. We used the Adam optimizer without pretrained weights on the encoder to ensure the model learns features specific to our dataset.

- **Loss Function and Training:** We employed binary cross-entropy as the loss function, ideal for the binary segmentation tasks. Training involved a batch size of 32 and was conducted over 50 epochs. The learning rate was managed by a ReduceLROnPlateau callback, which reduced the learning rate by a factor of 0.5 if no improvement was seen in the validation loss for four epochs, with a minimum learning rate set to 1×10^{-6}. Additionally, early stopping was implemented to terminate training if the validation loss did not improve for 25 consecutive epochs, helping prevent overfitting.

- **Callbacks and Checkpoints:** Model checkpoints were used to save the weights of the best performing model based on validation loss. This approach ensures that the model can be restored to its most effective state post-training.

Classification

- **Models Tested:** Post-segmentation, the isolated lung regions were processed through several classification models to detect the presence of COVID-19, viral pneumonia, or normal conditions. The models tested included ResNet50, CoAtNet0 (proposed model), XceptionNet, and InceptionResNetV2.

- **Selection:** CoAtNet[7] emerged as the most effective model, offering an optimal balance of depth and breadth in feature analysis. unlike existing methods that incorporate only CNNs or ViTs, we have adapted CoAtNet owing to its hybrid convolutional and transformer architecture. This combination harnesses the local feature extraction prowess of CNNs with the global data processing ability of ViTs, enabling a more comprehensive analysis of chest X-ray (CXR) images. CoAtNet's ability to analyze features at multiple scales also ensures that no critical information is missed. It provided the most accurate classification results among the models tested.

- **Data Augmentation:** For the classification stage, data augmentation techniques were applied to the training dataset to enhance model robustness and help prevent overfitting. These included random rotations up to 20 degrees, width and height shifts up to 20%, shear transformations up to 20%, zoom up to 20%, and nearest fill mode for newly introduced pixels.

- **Hyperparameters:** The model was set up with an input shape of 256×256 pixels for grayscale images. The model was compiled with the Adam optimizer, starting with a learning rate of 1×10^{-3}. We used sparse categorical crossentropy for the loss function due to its efficacy in handling multi-class classification problems. Training was set for 50 epochs with a batch size of 32.

- **Callbacks and Checkpoints:** Early stopping was set to terminate training if the validation loss did not improve for 15 epochs, with the best weights restored at the end. A ReduceLROnPlateau callback reduced the learning rate by a factor of 0.2 if the validation loss did not improve for seven epochs, with a verbose flag for logging and a minimum learning rate set to 110^{-7}. A ModelCheckpoint was used to save the best model based on the highest validation sparse categorical accuracy.

- **Testing:** Post-training, the model was evaluated on a separate test set using the test generator, with results reported in terms of loss and accuracy.

Localization with Grad-CAM: After classifying the lung regions, Grad-CAM was employed to localize the specific areas within the lung that significantly influenced the classification decision. This step enhances the model's interpretability by providing visual explanations for the predictions, crucial for clinical assessments and further validation of the AI-driven diagnostic process.

Dataset Used

For our project, we are using the **COVID-19 Radiography Database** [10], a valuable resource developed by researchers from Qatar University, Doha, Qatar, the University of Dhaka, Bangladesh, and their international collaborators. This comprehensive database contains chest X-ray images for three distinct classes: COVID-19, normal, and viral pneumonia. Specifically, the dataset comprises 3616 images of COVID-19 positive cases, 10,192 images categorized as normal, and 1345 images identified as viral pneumonia. This extensive collection allows us to train our diagnostic models effectively, ensuring robust performance in identifying and classifying these conditions.

Model Architecture

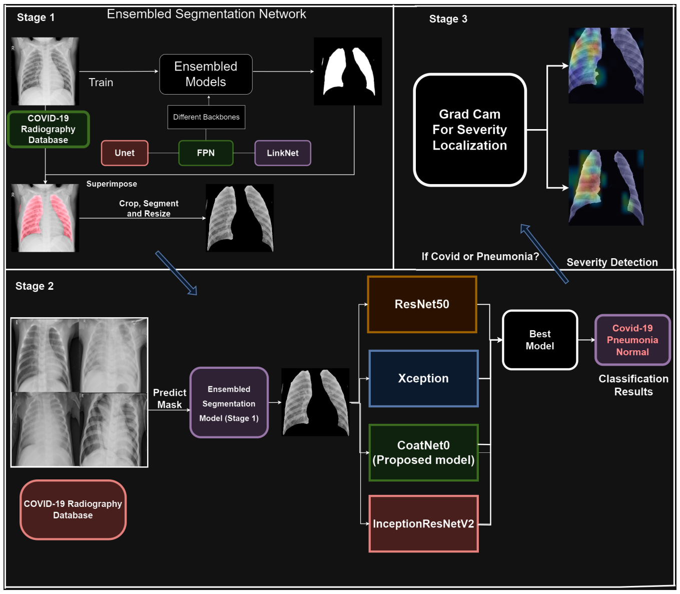

Figure 1: Model Architecture

Results

The integrated approach of segmentation using uNet, classification using CoAtNet, and localization using Grad-CAM significantly enhanced diagnostic accuracy. The model achieved an accuracy improvement of over 3% compared to baseline models and effectively localized the disease regions within the lung segments, providing critical insights for clinical diagnosis.

Results of Segmentation

In the segmentation stage, multiple models like U-Net, FPN, and LinkNet, each paired with various backbones [9] such as VGG16, SEResNext50, InceptionResNetV2, MobileNetV2, and EfficientNetB2 were tested. The U-Net model with the InceptionResNetV2 backbone proved to be the most effective, achieving the highest accuracy and efficiency in segmenting lung regions from CXR images.

Backbone	**binary_accuracy**	**dice_coef**	**iou_score**
Mobilenetv2	0.7585	0.5252	0.7511
seresnext50	0.9886	0.9757	0.9941
inceptionresnetv2	0.9906	0.9794	0.995
efficientnetb2	0.9894	0.9769	0.9945
vgg16	0.9282	0.7714	0.9403

Table 1: Performance Metrics of U-Net Models

The excellent performance of the U-Net model with the InceptionResNetV2 backbone can be explained by its strong ability to extract features, benefiting from the detailed and complex structure of the InceptionResNetV2 architecture. This backbone improves the model's capacity to identify detailed aspects within the CXR images, which is essential for accurately outlining lung borders and other minor pathological details. The detailed performance metrics of each segmentation model are presented in the accompanying table, emphasizing the superior capabilities of the chosen U-Net model.

Backbone	**binary_accuracy**	**dice_coef**	**iou_score**
Mobilenetv2	0.6167	0.4094	0.7271
seresnext50	0.9967	0.9925	0.9981
inceptionresnetv2	0.998	0.9946	0.9985
efficientnetb2	0.9905	0.9794	0.995
vgg16	0.989	0.969	0.9907

Table 2: Performance Metrics of LinkNet Models

The FPN and LinkNet models with the MobileNetV2 backbone did not perform as well in segmenting lung regions from CXR images. This lesser performance could be because MobileNetV2 is designed to be fast and efficient rather than focusing on capturing detailed features. MobileNetV2 uses a type of processing that is good for quick computations but not for identifying complex patterns in the images. This makes it difficult for these models to detect the small, important details in lung structures, which are necessary for accurate medical image analysis.

Backbone	**binary_accuracy**	**dice_coef**	**iou_score**
Mobilenetv2	0.6587	0.4794	0.7374
seresnext50	0.998	0.9946	0.9985
inceptionresnetv2	0.9913	0.9818	0.9955
efficientnetb2	0.9920	0.9826	0.9957
vgg16	0.9911	0.972	0.9917

Table 3: Performance Metrics of FPN Models

Results of Classification

In the classification stage of our study, we evaluated four different models: CoAtNet, Xception, ResNet50, and InceptionResNetV2, focusing on their performance across validation and test datasets. The results clearly demonstrated that CoAtNet outperformed the other models, achieving the lowest validation loss of 0.159 and the highest validation accuracy of 95.8%, as well as the lowest test loss of 0.144 and the highest test accuracy of 95.49%.

This superior performance can be attributed to CoAtNet's innovative architecture that combines convolutional layers with transformer mechanisms, enabling it to effectively capture both local and global features within the chest X-ray images. This hybrid approach allows CoAtNet to achieve a better understanding of the intricate patterns and anomalies present in the images, leading to more accurate and reliable classification of COVID-19, viral pneumonia, and normal cases. The ability of CoAtNet to use the strengths of both CNNs and transformers effectively addresses the limitations observed in the other models, making it the most suitable choice for this application.

Model	Val Loss	Val Acc.	Test Loss	Test Acc.
CoAtNet0	**0.159**	**95.8%**	**0.144**	**95.49%**
Xception	0.182	95.63%	0.220	93.80%
ResNet50	0.182	93.11%	0.220	92.64%
InceptionResNet50	0.193	94.81%	0.197	93.87%

Table 4: Classification Models Performance

Localisation Visualization

Figure 2 illustrates the localization maps generated by Grad-CAM, highlighting the areas focused on by the classification model (CoatNet). These maps clearly show that our network targets the relevant areas of the lungs for each disease category. The ability of our model to localize these regions, as shown in Figure 2, underscores its potential utility in clinical settings, where efficient disease localization and annotation are crucial yet labor-intensive tasks.

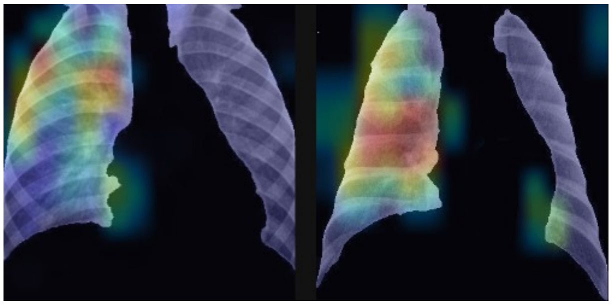

Figure 2: Localisation using Grad-CAM

Future Work

Future work will involve collaboration with radiologists to verify the regions identified by our model, ensuring their clinical relevance. This step is critical for assessing the practical effectiveness of our localization without prior comparative data. This verification will also help refine the model's performance, potentially leading to more accurate diagnostic tools that can support clinical decision-making and improve patient outcomes.

Conclusion

This project successfully develops a comprehensive framework for the detection and classification of respiratory diseases from CXR images. By integrating precise lung segmentation, robust classification, and effective localization techniques, the model achieves high accuracy and reliability, making it a valuable tool for clinical diagnostics.

Code Repository

The source code and datasets used in this project are available at **GitHub Repo**.

References

1. M. H. Nguyen and K. N. Quang, “A Study of Vision Transformer for Lung Diseases Classification,” in *Proceedings of 2022 6th International Conference on Green Technology and Sustainable Development*, GTSD 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 116–121. doi: 10.1109/GTSD54989.2022.9989100.
2. A. Saood and I. Hatem, “COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet,” *BMC Med Imaging*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12880-020-00529-5.
3. E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, “Deep learning for chest X-ray analysis: A survey,” *Med Image Anal*, vol. 72, p. 102125, Aug. 2021, doi: 10.1016/J.MEDIA.2021.102125.
4. A. S. Panayides et al., “AI in Medical Imaging Informatics: Current Challenges and Future Directions,” *IEEE J Biomed Health Inform*, vol. 24, no. 7, p. 1837, Jul. 2020, doi: 10.1109/JBHI.2020.2991043.
5. S. Bharati, P. Podder, and M. R. H. Mondal, “Hybrid deep learning for detecting lung diseases from X-ray images,” *Inform Med Unlocked*, vol. 20, p. 100391, Jan. 2020, doi: 10.1016/J.IMU.2020.100391.
6. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Oct. 2016, doi: 10.1007/s11263-019-01228-7.
7. Z. Dai, H. Liu, Q. V. Le, and M. Tan, “CoAtNet: Marrying Convolution and Attention for All Data Sizes,” Jun. 2021, [Online]. Available: `http://arxiv.org/abs/2106.04803`
8. T. A. Pham and V. D. Hoang, “Chest X-ray image classification using transfer learning and hyperparameter customization for lung disease diagnosis,” *Journal of Information and Telecommunication*, 2024, doi: 10.1080/24751839.2024.2317509.
9. “qubvel/segmentation_models: Segmentation models with pretrained backbones. Keras and TensorFlow Keras.” Accessed: May 12, 2024. [Online]. Available: `https://github.com/qubvel/segmentation_models`
10. M. E. H. Chowdhury et al., “Can AI Help in Screening Viral and COVID-19 Pneumonia?,” *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.